Exploring Fixed Point in Image Editing: Theoretical Support and Convergence Optimization

Chen Hang[1]*, **Zhe Ma**[2]*, **Haoming Chen**[1], **Xuwei Fang**[3], **Weisheng Xie**[3],
Faming Fang[1]†, **Guixu Zhang**[1], **Hongbin Wang**[2]†
[1]East China Normal University, Shanghai, 200062 China
[2]Ant Group, Hangzhou, 310063 China
[3]Bestpay AI Lab, Shanghai, 200080 China

Abstract

In image editing, Denoising Diffusion Implicit Models (DDIM) inversion has become a widely adopted method and is extensively used in various image editing approaches. The core concept of DDIM inversion stems from the deterministic sampling technique of DDIM, which allows the DDIM process to be viewed as an Ordinary Differential Equation (ODE) process that is reversible. This enables the prediction of corresponding noise from a reference image, ensuring that the restored image from this noise remains consistent with the reference image. Image editing exploits this property by modifying the cross-attention between text and images to edit specific objects while preserving the remaining regions. However, in the DDIM inversion, using the $t-1$ time step to approximate the noise prediction at time step t introduces errors between the restored image and the reference image. Recent approaches have modeled each step of the DDIM inversion process as finding a fixed-point problem of an implicit function. This approach significantly mitigates the error in the restored image but lacks theoretical support regarding the existence of such fixed points. Therefore, this paper focuses on the study of fixed points in DDIM inversion and provides theoretical support. Based on the obtained theoretical insights, we further optimize the loss function for the convergence of fixed points in the original DDIM inversion, improving the visual quality of the edited image. Finally, we extend the fixed-point based image editing to the application of unsupervised image dehazing, introducing a novel text-based approach for unsupervised dehazing.

1 Introduction

Diffusion models have gained significant attention in recent years. Early versions of Denoising Diffusion Probabilistic Models (DDPMs) [1] and score-based generative models [2] were capable of generating high-quality images and even surpassed the performance of Generative Adversarial Networks (GANs) [3] at that time. However, the initial diffusion models required a large number of sampling steps during image generation, which was a major bottleneck of diffusion models. Subsequent research focused on reducing the number of sampling steps without compromising the quality of generated images [4]. During this period, DDIM [5] was proposed and widely used for its speed and flexibility in both deterministic and stochastic generation. When DDIM is set to deterministic sampling, it can be viewed as an ODE process that is reversible. This property enables realistic image editing, leading to the introduction of Prompt-to-Prompt (P2P) [6] editing. The core idea of P2P is to modify the cross-attention between text and images. For local edits, such as replacing

*Equal contribution.

†Corresponding Author, fmfang@cs.ecnu.edu.cn, hongbin.whb@antgroup.com

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

a cat with a dog while keeping the background unchanged, only the cross-attention corresponding to the cat needs to be modified. For enhancing certain effects, the cross-attention corresponding to the modifying word is reweighted. However, P2P still suffers from unreliable results in realistic image editing, as it exhibits inconsistencies even when directly reconstructing the reference image without any edits. The main cause of this issue lies in the errors introduced during DDIM inversion, which will be discussed later in this paper. To address this problem, some methods introduce auxiliary variables to mitigate the issue, such as null-text inversion (NTI) [7] for learning null-text embeddings. Exact diffusion inversion (EDICT) [8] uses two coupled vectors to invert each other in an alternating fashion, to achieve precise diffusion inversion. In recent methods, direct inversion (DI) [9] and inversion-free editing (InfEdit) [10] decouple the source and target branches to avoid optimizing DDIM inversion and achieve satisfactory results. However, these methods do not address the error issue in DDIM inversion theoretically. This issue can be formulated as solving fixed points of implicit functions, which has been mentioned in accelerated iterative diffusion inversion (AIDI) [11] and fixed-point inversion (FPI) [12]. However, AIDI and FPI only employ methods to solve fixed points of implicit functions without proving the existence of fixed points at each sampling step of DDIM inversion. Only FPI indirectly demonstrates the convergence of fixed-point loss through numerical experiments on the used dataset. Such numerical experiments cannot rigorously prove the existence of fixed points. Therefore, in this paper, we rigorously prove the existence and uniqueness of fixed points in DDIM inversion using the Banach fixed-point theorem [13]. Subsequently, based on this uniqueness, we identify flaws in the fixed-point loss used by AIDI and FPI and propose optimization. Finally, we explore the application of fixed-point based image editing in unsupervised image dehazing. In summary, our contributions in this paper are as follows:

- We theoretically prove that the Lipschitz constant in DDIM inversion is less than one. Based on the Banach fixed-point theorem, we establish the existence and uniqueness of fixed points, providing theoretical support for image editing methods that involve solving fixed points of implicit functions.
- Leveraging the uniqueness of fixed points, we demonstrate through theory and experimental cases that the existing methods suffer from flaws in the fixed-point loss and propose optimization.
- We extend the fixed-point based image editing approach to the task of unsupervised image dehazing and explore the feasibility of text-guided unsupervised dehazing through fixed-point based editing.

2 Preliminaries: DDPMs and DDIM Inversion

DDPMs model the probability distribution of a diffusion process by iteratively adding noise to a data sample until the data distribution becomes predominantly noise [1]. Subsequently, the data distribution is recovered from a random Gaussian noise through a reverse diffusion process. The diffusion process is typically regarded as a Markov chain starting from x_0 and obtaining $x_1, x_2, ..., x_T$ by adding noise. The specific process is outlined as follows:

$$q\left(x_{1:T} \mid x_0\right) := \prod_{t=1}^{T} q\left(x_t \mid x_{t-1}\right), \quad q\left(x_t \mid x_{t-1}\right) := \mathcal{N}\left(x_t; \sqrt{1-\beta_t} x_{t-1}, \beta_t I\right), \tag{1}$$

where schedule $\beta_0, \beta_1, ..., \beta_T \in (0,1)$. x_t can be expressed as a linear combination of x_0 and Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \quad \alpha_t := 1-\beta_t, \quad \bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s, \tag{2}$$

In DDPMs, both the diffusion process and the reverse diffusion process are Markov chains, which require a large number of sampling steps during image generation. However, DDIM introduces a method to accelerate sampling by decoupling the reverse diffusion process from the Markov chain, making it feasible to perform skip-step sampling and reducing the number of sampling steps [5]. Simultaneously, the introduction of Stable Diffusion (SD) [14] alleviates the resource requirements for training diffusion models and improves inference speed. SD involves compressing the image into a latent space for diffusion instead of operating in the original pixel-level space. As a result, the reverse diffusion process can be expressed as follows:

$$z_{t-1} = \sqrt{\frac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}} z_t + \left(\sqrt{1-\bar{\alpha}_{t-1}} - \sqrt{\frac{(1-\bar{\alpha}_t)\bar{\alpha}_{t-1}}{\bar{\alpha}_t}} \right) \epsilon_\theta\left(z_t, t, p\right), \tag{3}$$

where ϵ_θ represents a network that predicts added noise, z_t denotes the compressed representation of x_t in the latent space, t represents the time step, and p represents the encoding of a condition text.

Now let's focus on DDIM inversion. DDIM inversion seeks to find a noise z_T given the latent space representation z_0 corresponding to a given x_0 and the associated textual prompt p. This allows for the autoregressive reconstruction of z_0. The process of DDIM inversion can be rewritten by Equation 3 as follows:

$$\begin{aligned} z_t &= \sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} z_{t-1} + \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\frac{(1-\bar{\alpha}_{t-1})\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} \right) \epsilon_\theta\left(z_t, t, p\right) \\ &\approx \sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} z_{t-1} + \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\frac{(1-\bar{\alpha}_{t-1})\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} \right) \epsilon_\theta\left(z_{t-1}, t, p\right) \end{aligned} \tag{4}$$

Therefore, we can invert z_0 back to z_T using Equation 4, allowing us to reconstruct z_0 using the reverse diffusion process. However, it can be observed that in Equation 4, z_t is directly approximated by z_{t-1}, which inevitably introduces errors that accumulate over time. This inconsistency between the forward and reverse processes of the diffusion leads to poor quality in image reconstruction and editing [7, 8].

3 Theoretical Support for Fixed Point in DDIM Inversion

From the first equation of Equation 4, it is evident that finding z_t can be regarded as solving a fixed-point problem for the implicit function with respect to $\epsilon_\theta\left(z_t, t, p\right)$. Therefore, we can consider the right-hand side of the equation as a function with z_t as a variable:

$$f(z_t) = \sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} z_{t-1} + \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\frac{(1-\bar{\alpha}_{t-1})\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} \right) \epsilon_\theta\left(z_t, t, p\right), \tag{5}$$

Then, to find $f(z_t) = z_t$, we can use z_{t-1} as the starting point for iteration. The iteration process can be expressed as follows:

$$z_t^0 = z_{t-1}, \quad z_t^N = z_t, \quad z_t^{i+1} = f\left(z_t^i\right), \quad (i = 0, 1, ..., N) \tag{6}$$

After multiple iterations, the fixed point of Equation 5 can be found. Now, we focus on the theoretical analysis of whether a fixed point exists in Equation 5. Let z_t^i and z_t^j represent any two points in the function $f(z_t)$ at time step t, where i and j, with j greater than i, are within the range of 0 to N. We can obtain the following:

$$\begin{aligned} ||f(z_t^i) - f(z_t^j)|| &= \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\frac{(1-\bar{\alpha}_{t-1})\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} \right) ||\epsilon_\theta(z_t^i, t, p) - \epsilon_\theta(z_t^j, t, p)|| \\ &\leq \sqrt{1-\bar{\alpha}_t} ||\epsilon_\theta(z_t^i, t, p) - \epsilon_\theta(z_t^j, t, p)|| \end{aligned} \tag{7}$$

where $|| \cdot ||$ is L_2 norm. We know that in the reverse diffusion process, when predicting the z value at time step $t-1$ given the z value at time step t, an initial rough estimation of z_0 is made. This rough estimation of z_0 introduces some errors compared to the final obtained value of z_0. However, the overall shape of the image is already quite similar, and as time t decreases, this rough estimation of z_0 gradually approaches the final obtained z_0. This rough estimation of z_0 can be represented as follows:

$$z_0^i = \left(z_t^i - \sqrt{1-\bar{\alpha}_t} \epsilon_\theta(z_t^i, t, p)\right) / \sqrt{\bar{\alpha}_t}, \quad z_0^j = \left(z_t^j - \sqrt{1-\bar{\alpha}_t} \epsilon_\theta(z_t^j, t, p)\right) / \sqrt{\bar{\alpha}_t}, \tag{8}$$

where z_0^i and z_0^j represent the predictions of time step zero. Although z_0^i and z_0^j obtained from z_t^i and z_t^j are very close, there might be some errors when t is not close to zero. Let's denote this error as ε_0. Additionally, let's define the difference between z_t^i and z_t^j as ε_t. With Equation 8, we can derive the following relationship:

$$z_t^i - z_t^j = \sqrt{\bar{\alpha}_t}\varepsilon_0 + \sqrt{1-\bar{\alpha}_t}\left(\epsilon_\theta(z_t^i, t, p) - \epsilon_\theta(z_t^j, t, p)\right), \tag{9}$$

Due to the fact that different initial values at time step t can lead to similar z_0, i.e., any position at time step t can converge to a small range defined by the rough estimation of z_0, we can obtain $|\varepsilon_0| = k|\varepsilon_t|$, where $0 < k < 1$. Moreover, from the perspective of ODE, the paths from z_t^i to z_0^i and from z_t^j to z_0^j can be viewed as two ODE trajectories with initial values of z_t^i and z_t^j, respectively, or as two integral curves of ODE given the initial values z_t^i and z_t^j. These integral curves non-tangent and non-intersecting [15]. Therefore, we can conclude that ε_0 and ε_t have the same sign. We can also prove this by contradiction: if ε_0 and ε_t have opposite signs, then the two ODE trajectories would have an intersection point. From that intersection point, there would exist two different paths in both forward and backward directions, which contradicts the deterministic sampling of DDIM. Consequently, we can deduce $\varepsilon_0 = k\varepsilon_t$, where $0 < k < 1$. By substituting this into Equation 9, we obtain:

$$\begin{aligned} z_t^i - z_t^j &= k\sqrt{\bar{\alpha}_t}\varepsilon_t + \sqrt{1-\bar{\alpha}_t}\left(\epsilon_\theta(z_t^i, t, p) - \epsilon_\theta(z_t^j, t, p)\right) \\ &= k\sqrt{\bar{\alpha}_t}\left(z_t^i - z_t^j\right) + \sqrt{1-\bar{\alpha}_t}\left(\epsilon_\theta(z_t^i, t, p) - \epsilon_\theta(z_t^j, t, p)\right) \end{aligned} \tag{10}$$

By rearranging the Equation 10 and substituting it into Equation 7, we obtain:

$$||f(z_t^i) - f(z_t^j)|| \leq \left(1 - k\sqrt{\bar{\alpha}_t}\right)||z_t^i - z_t^j||, \tag{11}$$

From Equation 2, it is known that $\bar{\alpha}_t \in (0, 1)$. Therefore, it can be inferred that $(1 - k\sqrt{\bar{\alpha}_t}) \in (0, 1)$. This proves that the implicit function $f(z_t)$ satisfies the contraction mapping condition. By further applying the Banach fixed-point theorem [13], we can conclude that the fixed point of $f(z_t)$ exists and is unique.

4 Convergence Optimization for Fixed Point in DDIM Inversion

In AIDI [11] and FPI [12], to determine the convergence of the fixed point obtained through iterative processes in the implicit function $f(z_t)$, the criterion used is $f(z_t) - z_t$. In FPI, numerical experiments were conducted on the given dataset, and it was concluded that convergence is achieved within $2 \sim 3$ iterations. Furthermore, in Section 3, we have proved the uniqueness of the fixed point. Therefore, at any position with a time step t, convergence can be reached towards the same point. We express this using Equation 11 as follows:

$$\begin{aligned} ||f(z_t^i) - f(z_t^j)|| &\leq \left(1 - k\sqrt{\bar{\alpha}_t}\right)||z_t^i - z_t^j||, \\ ||f(\hat{z}_t^i) - f(\hat{z}_t^j)|| &\leq \left(1 - k\sqrt{\bar{\alpha}_t}\right)||\hat{z}_t^i - \hat{z}_t^j||, \\ ||f(z_t^j) - f(\hat{z}_t^j)|| &\leq \left(1 - k\sqrt{\bar{\alpha}_t}\right)||z_t^j - \hat{z}_t^j||, \end{aligned} \tag{12}$$

where $\hat{z}_t^i = z_t^i + \epsilon, \epsilon \sim \mathcal{N}(0, I)$, and $\hat{z}_t^j$ is the result obtained by iterating $\hat{z}_t^i$ in the implicit function $f(z_t)$. It can also be considered that $\hat{z}_t^i$ is a new path distinct from z_t^i. Theoretically, the three cases of Equation 12 have the same convergence rate, which means they have the same Lipschitz constant. However, in practical computations, it is challenging for two different paths to converge exactly to the same fixed point, resulting in numerical errors. Consequently, this leads to the persistence of numerical discrepancies in the third inequality of Equation 12 even at the end of the iteration.

$$||f(z_t^N) - f(\hat{z}_t^N)|| > \varepsilon, \quad \varepsilon > 0, \tag{13}$$

Incorporating the error into the third inequality of Equation 12 for correction, we have:

$$\begin{aligned} ||f(z_t^j) - f(\hat{z}_t^j)|| &\leq \left(1 - k\sqrt{\bar{\alpha}_t}\right)||z_t^j - \hat{z}_t^j|| + \varepsilon \\ &\leq \left(1 - k\sqrt{\bar{\alpha}_t} + \delta\right)||z_t^j - \hat{z}_t^j||, \quad \delta > 0 \end{aligned} \tag{14}$$

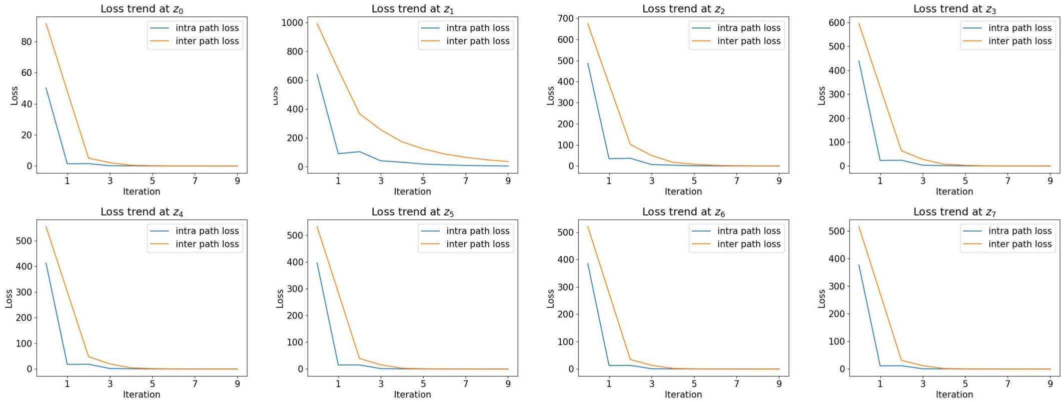


Figure 1: In the diffusion model with a step size of 50 steps, we selected the trend of fixed-point loss for the first 8 steps, while the trends for the remaining steps were consistent and are provided in the appendix A.1. In each step, we performed 10 iterations. The blue intra path loss represents $f(z_t) - z_t$ used in AIDI and FPI, while the orange inter path loss represents our optimized $f(z_t) - f(\hat{z}_t)$. The trend of the intra path loss aligns with the findings reported in FPI, converging within $2 \sim 3$ iterations. However, the inter path loss exhibits noticeable lag. Based on the proven uniqueness of fixed points in this paper, the convergence of the intra path loss does not indicate the convergence of fixed points. Instead, attention should be paid to the inter path loss to assess the convergence of fixed points.

It is not difficult to observe that if there are multiple paths and each path only focuses on the variation of the loss within the path, they will have the same convergence rate. However, when considering the loss between paths, there will be a noticeable lag in the convergence rate, as shown in Figure 1. The loss function $f(z_t) - z_t$ used in AIDI and FPI precisely corresponds to the case of solely focusing on the variation within paths. As a result, the fixed points obtained when the $f(z_t) - z_t$ loss converges may not be the optimal points. To optimize the original criterion for judging the convergence of fixed points, we modify it to $f(z_t) - f(\hat{z}_t)$ and define the convergence of the $f(z_t) - f(\hat{z}_t)$ loss as the convergence of fixed points. Through experimentation, we have also confirmed that the convergence of fixed points actually lags behind what is mentioned in FPI, as illustrated in Figure 1. In order to visually demonstrate the process of fixed point convergence, we randomly select two points from z_t as the axes. Subsequently, multiple paths are sampled by applying $\hat{z}_t^i = z_t^i + \epsilon, \epsilon \sim \mathcal{N}(0, I)$. The results obtained from multiple iterations are plotted as trajectory lines, as shown in Figure 2.

By optimizing the fixed point convergence criterion and the fixed-point loss, the computed fixed points will be more accurate. To further improve the precision of fixed points, it is natural, as suggested by Figure 2, to consider selecting the final points from multiple paths, finding the cluster centers of these points, and using them as the results of the fixed points. However, this approach would increase the computational burden. Therefore, in subsequent experiments, we only consider the results obtained when the inter path loss converges.

5 Generalization of Fixed Point in DDIM Inversion to Image Dehazing

Image dehazing, as a subtask of low-level vision, primarily aims to remove the haze degradation from hazy images to recover clean images, thereby assisting high-level visual tasks [16, 17] for improved performance. Hazy images are typically modeled as follows [18]:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{15}$$

where $I(x)$ is hazy image, $J(x)$ is the real scene to be recovered. $t(x) = e^{-kd(x)}$ is the medium transmission, where $d(x)$ is depth and k is scattering coefficient of the atmosphere. A is the global atmospheric light.

Existing unsupervised image dehazing methods can be broadly categorized into learning-based and non-learning-based approaches. Learning-based methods primarily utilize unpaired images [19, 20, 21] to train networks for unsupervised dehazing, while non-learning-based methods rely on image priors such as dark channel prior [22] and rank-one prior [23]. With the development of visual

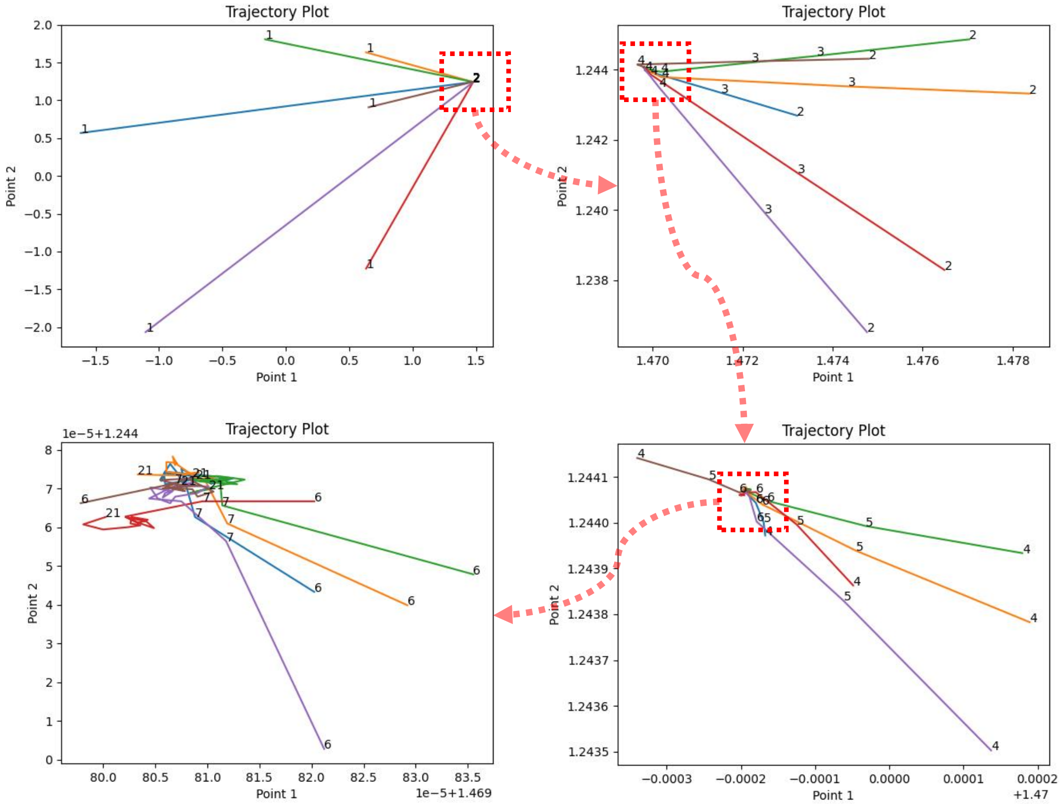


Figure 2: The axes were randomly chosen from two points in z_t. The total number of iterations to compute the fixed points was 21, and the corresponding step numbers were labeled in each path in each subgraph. It can be observed that the inter path loss starts to converge around the 7th step and beyond, indicating that the convergence of fixed points requires approximately 7 iterations.

language models, some supervised multimodal dehazing approaches have emerged [24]. However, there are few reports on unsupervised multimodal dehazing. This paper mainly explores image-editing-based unsupervised dehazing, where the main principle is to achieve dehazing effects through cross-attention replacement of corresponding to haze. Since the haze is replaced with null text, the null text optimization from NTI is necessary [7]. Otherwise, as the class-free guidance coefficient increases, the recovered image may collapse. The DDIM Inversion used in NTI does not involve fixed point correction. Therefore, in subsequent experiments, we will show that the use of fixed-point optimized NTI can effectively alleviate the image collapse issue in dehazed images.

6 Experiments

This section will be divided into two parts. The first part focuses on image editing experiments, aiming to highlight that modifying the convergence criterion of fixed points leads to better visual quality of the edited image compared to previous approaches. Additional techniques have also been introduced in AIDI and FPI to further enhance the visual effects of image editing. However, to avoid interference, we only reproduced their fixed point code. For the dataset, we selected PIE-Bench [9] in the context of image editing, which consists of 700 images featuring 10 distinct editing types.

The second part comprises experiments on unsupervised image dehazing. Our experiments are based on the NTI approach, incorporating fixed points to alleviate the issue of image collapse in NTI based dehazed images. We directly utilized the real-world scenes provided in the RESIDE [25] dataset for our experiments. Finally, both parts of the experiment employed identical configurations, comprising a sampling step of 50 and a class-free guidance coefficient of 7.5. In the context of image dehazing, to preclude any potential disruption of haze-free regions, no additional text descriptions were introduced. The original prompt was solely 'haze', while the target prompt was designated as '$\emptyset$'.

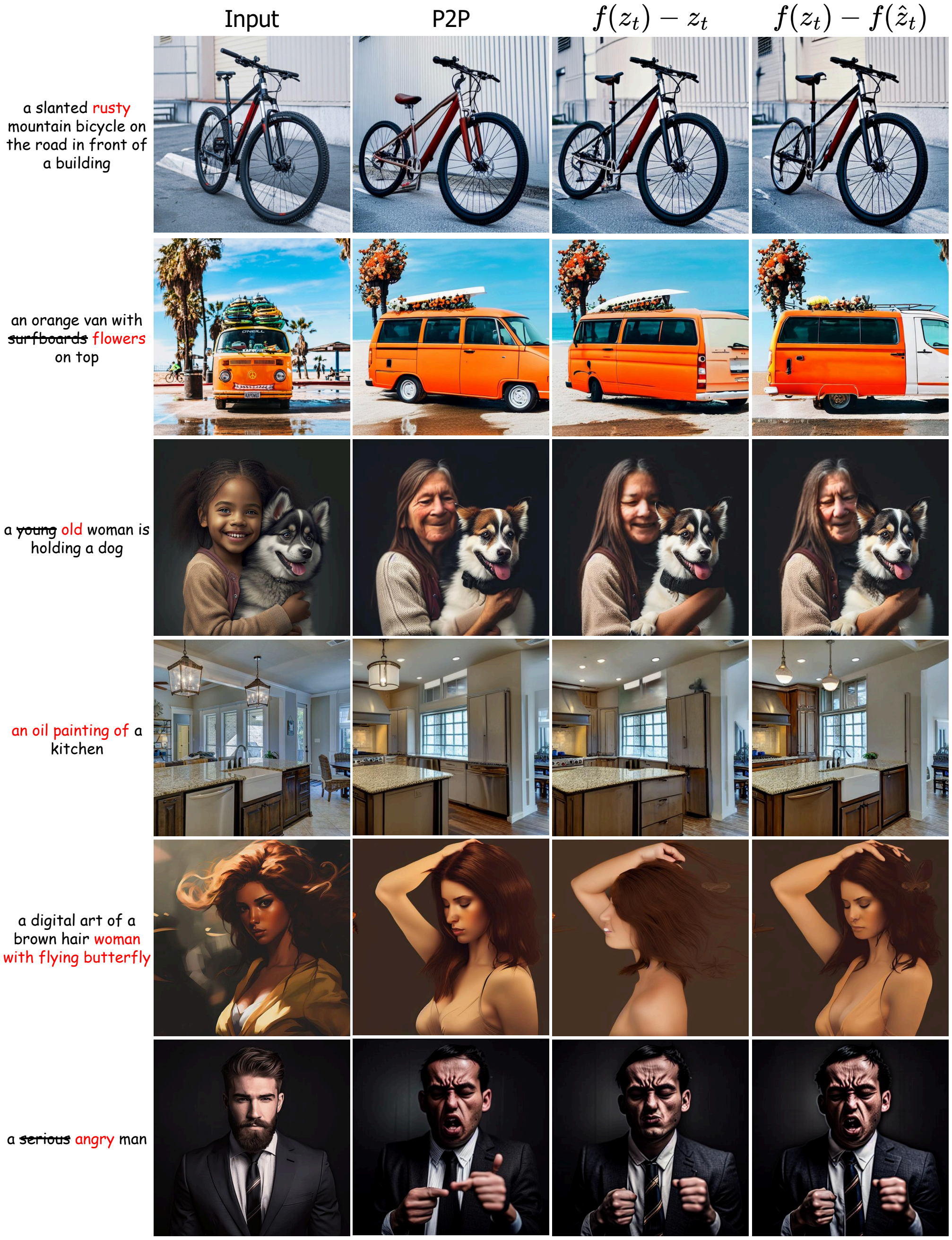

Input
P2P
$f(z_t) - z_t$
$f(z_t) - f(\hat{z}_t)$
a slanted rusty mountain bicycle on the road in front of a building
an orange van with ~~surfboards~~ flowers on top
a ~~young~~ old woman is holding a dog
an oil painting of a kitchen
a digital art of a brown hair woman with flying butterfly
a ~~serious~~ angry man

Figure 3: $f(z_t) - z_t$ denotes the intra path loss, while $f(z_t) - f(\hat{z}_t)$ signifies the optimized inter path loss. In the prompt, red signifies additions, whereas strikethrough indicates deletions. More results can be found in appendix A.2.

6.1 Image Editing

In the experiments on image editing, we primarily focused on the P2P and its variations: P2P with fixed points and P2P with improved convergence of fixed points. P2P serves as a baseline to demonstrate the effectiveness of fixed points. The proposed optimization for fixed point convergence is compared to the original fixed-point method to showcase the improved visual quality of the edited images. We evaluate the effectiveness of our proposed optimization from two perspectives. Firstly, we directly compare the imaging results of image editing, as shown in Figure 3 and Table 1. In

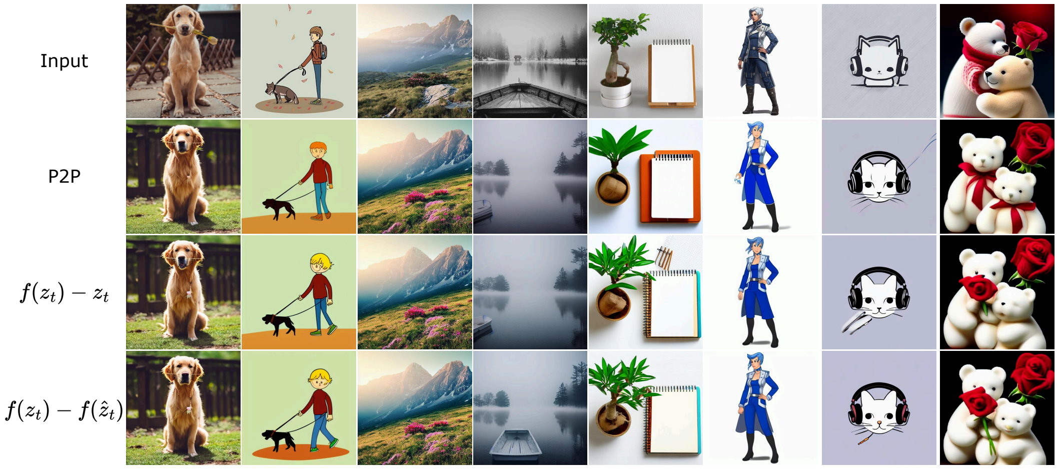


Figure 4: The autoregressive reconstruction result derived from noise subjected to DDIM inversion.

Table 1: Performance results for image editing and reconstruction, with evaluation metrics drawn from PIE-Bench [9], are presented across seven indicators spanning three dimensions. **Bold** indicate the best results.

Method	Structure	Background Preservation				CLIP Similariy	
	Distance$_\times 10^3$ ↓	PSNR ↑	LPIPS$_\times 10^3$ ↓	MSE$_\times 10^4$ ↓	SSIM$_\times 10^2$ ↑	Whole ↑	Edited ↑
	Image Editing / Image Reconstruction						
P2P	69.953 / 70.753	17.87 / 17.73	208.89 / 210.78	219.88 / 225.78	71.14 / 70.91	25.01 / 23.69	22.44 / 21.29
$f(z_t) - z_t$	69.521 / 70.495	17.83 / 17.73	208.98 / 210.76	221.48 / 225.60	71.14 / 70.91	25.19 / 23.66	22.53 / 21.33
$f(z_t) - f(\hat{z}_t)$	**69.190 / 70.046**	**17.87 / 17.74**	**208.88 / 210.71**	**219.85 / 225.50**	**71.15 / 70.91**	**25.31 / 23.80**	**22.58 / 21.37**

Figure 3, we selected some results that exhibit noticeable differences before and after optimization. It can be observed that optimizing the convergence of fixed points leads to improved performance and robustness. Secondly, we also evaluate the three methods in terms of autoregressive reconstruction of the images, as shown in Figure 4 and Table 1. It can be observed that using the optimized fixed points yields reconstructed images that have higher similarity to the input images.

6.2 Image Dehazing

In the unsupervised image dehazing experiments, we primarily compared the NTI approach with NTI incorporating fixed points. From the experimental results depicted in Figure 5 and Table 2, it can be observed that the inclusion of fixed points in NTI effectively mitigates the image collapse issue in certain dehazed images. As mentioned earlier, in unsupervised dehazing, where the attention for haze is replaced with null text attention, not using NTI would result in image collapse when the class-free guidance coefficient increases. Therefore, we also included examples of image collapse in Figure 5.

Table 2: Results of dehazing in real-world scenes, evaluation was conducted utilizing Non-Reference Image Quality Assessment (NR-IQA). **Bold** indicate the best results.

Method	NR-IQA	
	BRISQUE ↓	NIQE ↓
NTI	27.5879	4.4499
NTI w fixed point	**24.1297**	**3.4777**
fixed point w/o NTI	37.1827	7.7027

7 Limitations, Prospects and Impact

Although the fixed-point loss achieves better visual quality after optimization, there is still room for improvement. Strategies proposed in AIDI and FPI can be employed in image editing to further enhance the consistency between edited images and reference images. However, in this study, we did not reproduce the mentioned strategies in order to avoid interference with the optimization of the fixed-point. We now primarily discuss the limitations of unsupervised image dehazing and other image restoration tasks. In unsupervised dehazing, although some haze can be effectively removed

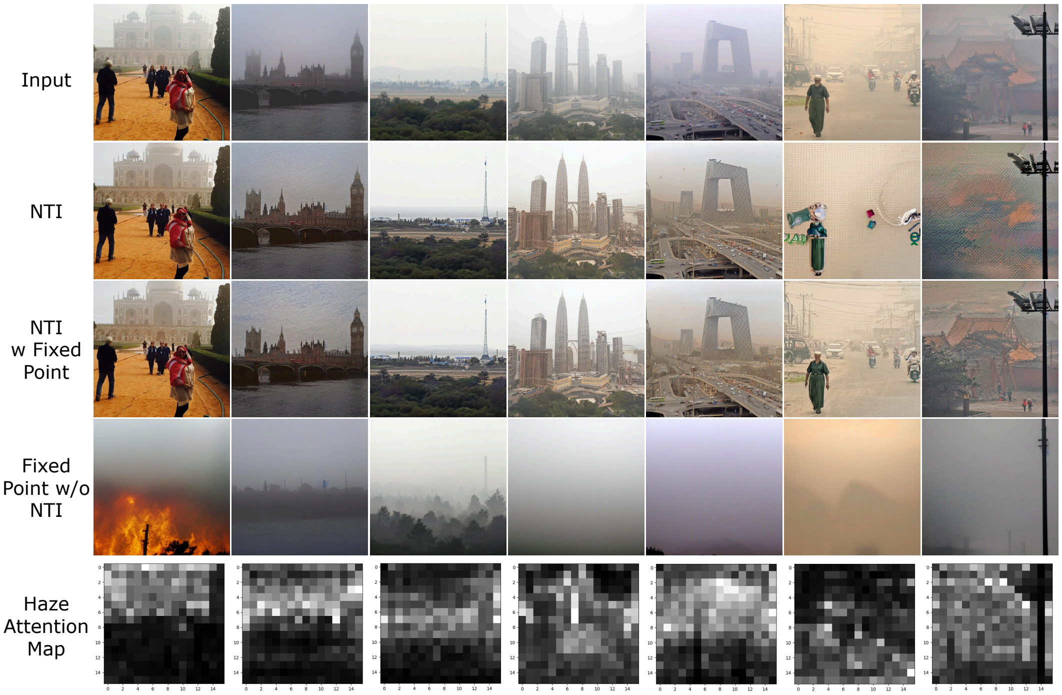


Figure 5: Imaging results of image dehazing under real-world scenes in RESIDE dataset.

from images, there are still issues with artifacts. For example, in the middle column of Figure 5, some artifacts can be observed. We believe that one reason is the imprecise attention map corresponding to the haze. We present the matched attention map in last line from Figure 5. It can be seen that the attention map corresponding to the haze semantics only provides a rough match, far from pixel-level alignment. After analyzing the process of attention map generation, we found that the attention map mainly resides in the deeper layers of the Unet. This implies that pixel-level fine-grained alignment cannot be achieved, which also hinders the direct transfer of image editing to some fine-grained image restoration tasks such as image deraining.

We have summarized some experiences in attempting to improve the performance of unsupervised image dehazing. A more straightforward approach is to fine-tune the dehazing task using existing mature techniques such as Lora [26] and Adapter [27], enabling the model to match more accurate haze attention maps. Another strategy that does not require training is to introduce image priors to weight the haze attention map. Since the attention map is primarily obtained by summing and averaging multiple layers of cross-attention, the priors need to be weighted into each layer of cross-attention. However, the cross-attention at different layers may have little correlation with the final haze attention map in terms of morphology. We attempted to directly weight the dark channel prior [22] and depth estimation prior [28] into each layer of cross-attention and found that this direct weighting method can improve the restoration results of some images but may also cause some images to collapse (see appendix A.3). Therefore, designing a weighted algorithm specifically for priors is necessary. In addition, similar to other generative models [29, 30, 31], the ability of image editing can be exploited to generate deceptive and harmful content, and the fixed-point based image editing may potentially exacerbate the negative impact of deep generative models for malicious purposes.

8 Conclusion and Future Work

This paper fills the theoretical gap in the fixed-point theory of DDIM inversion for image editing, providing theoretical support for image editing methods that employ fixed points. Additionally, based on the unique conclusion regarding the existence of fixed points obtained in this paper, we optimize the convergence criterion for fixed points and propose the inter path loss as a measure for assessing fixed-point convergence, supported by both theoretical analysis and experimental results. Through experiments conducted on PIE-Bench, we demonstrate improved visual quality after optimization. Finally, we extend the application of fixed-point based image editing to unsupervised image dehazing and analyze the limitations and prospects of image restoration tasks. In our future works, we will delve further into fixed-point computations and unsupervised semantic based image restoration. Regarding fixed points, our exploration will focus on reducing computational overhead associated with their calculation and enhancing the precision. Concerning unsupervised image restoration, we will primarily examine ways to improve the corrective impact of priors on imprecise semantic attention maps, thereby achieving superior image restoration outcomes.

9 Acknowledgements

This work was supported by the National Key R&D Program of China (2022ZD0161800), and the National Natural Science Foundation of China under Grant 62271203. Besides, this work were supported by Ant Group and Bestpay AI Lab.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[2] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[4] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023.

[5] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

[6] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*, 2022.

[7] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6038–6047, 2023.

[8] Bram Wallace, Akash Gokul, and Nikhil Naik. Edict: Exact diffusion inversion via coupled transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22532–22541, 2023.

[9] Xuan Ju, Ailing Zeng, Yuxuan Bian, Shaoteng Liu, and Qiang Xu. Direct inversion: Boosting diffusion-based editing with 3 lines of code. *arXiv preprint arXiv:2310.01506*, 2023.

[10] Sihan Xu, Yidong Huang, Jiayi Pan, Ziqiao Ma, and Joyce Chai. Inversion-free image editing with natural language. *arXiv preprint arXiv:2312.04965*, 2023.

[11] Zhihong Pan, Riccardo Gherardi, Xiufeng Xie, and Stephen Huang. Effective real image editing with accelerated iterative diffusion inversion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15912–15921, 2023.

[12] Barak Meiri, Dvir Samuel, Nir Darshan, Gal Chechik, Shai Avidan, and Rami Ben-Ari. Fixed-point inversion for text-to-image diffusion models. *arXiv preprint arXiv:2312.12540*, 2023.

[13] Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta mathematicae*, 3(1):133–181, 1922.

[14] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[15] Witold Hurewicz. *Lectures on ordinary differential equations*. Courier Corporation, 1990.

[16] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[17] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[18] Srinivasa G Narasimhan and Shree K Nayar. Chromatic framework for vision in bad weather. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, pages 598–605. IEEE, 2000.

[19] Yang Yang, Chaoyue Wang, Xiaojie Guo, and Dacheng Tao. Robust unpaired image dehazing via density and depth decomposition. *International Journal of Computer Vision*, pages 1–21, 2023.

[20] Bosheng Ding, Ruiheng Zhang, Lixin Xu, Guanyu Liu, Shuo Yang, Yumeng Liu, and Qi Zhang. U 2 d 2 net: Unsupervised unified image dehazing and denoising network for single hazy image enhancement. *IEEE Transactions on Multimedia*, 2023.

[21] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2037–2046, 2022.

[22] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[23] Jun Liu, Ryan Wen Liu, Jianing Sun, and Tieyong Zeng. Rank-one prior: Real-time scene recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[24] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Photo-realistic image restoration in the wild with controlled vision-language models. *arXiv preprint arXiv:2404.09732*, 2024.

[25] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.

[26] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[27] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pages 2790–2799. PMLR, 2019.

[28] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. *arXiv preprint arXiv:2401.10891*, 2024.

[29] Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models. *Advances in Neural Information Processing Systems*, 35:16782–16795, 2022.

[30] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022.

[31] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. *Advances in Neural Information Processing Systems*, 35:14715–14728, 2022.

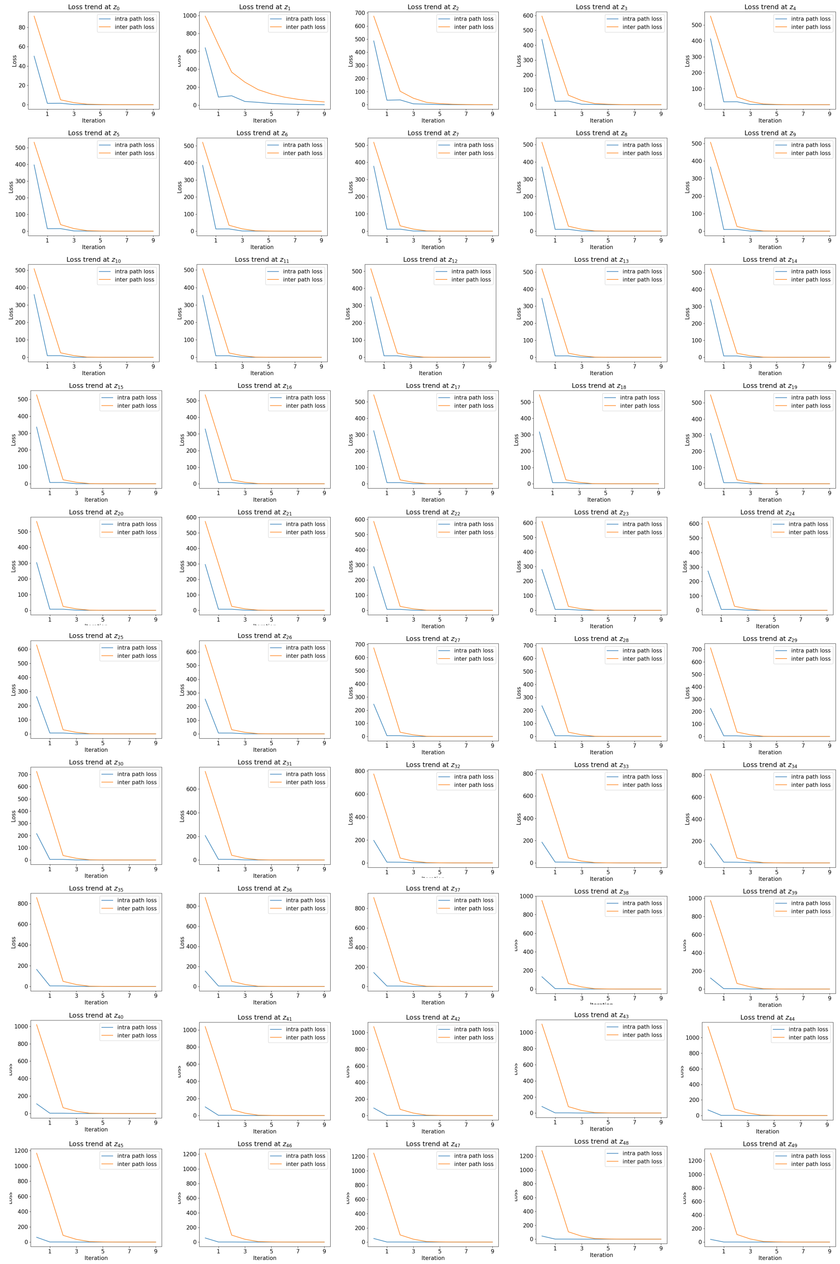


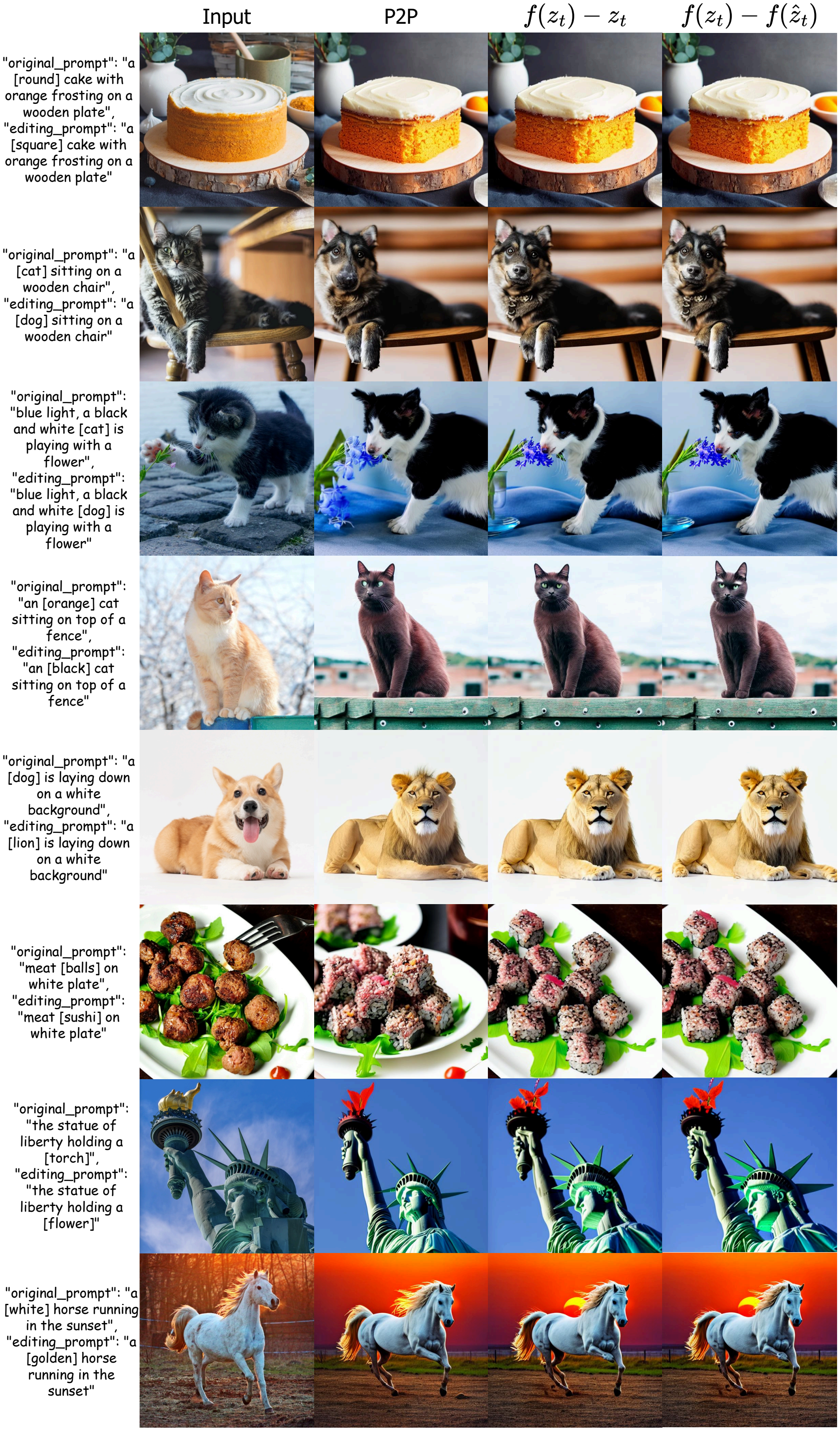
Figure 6: The trend of fixed-point loss for 50 steps.

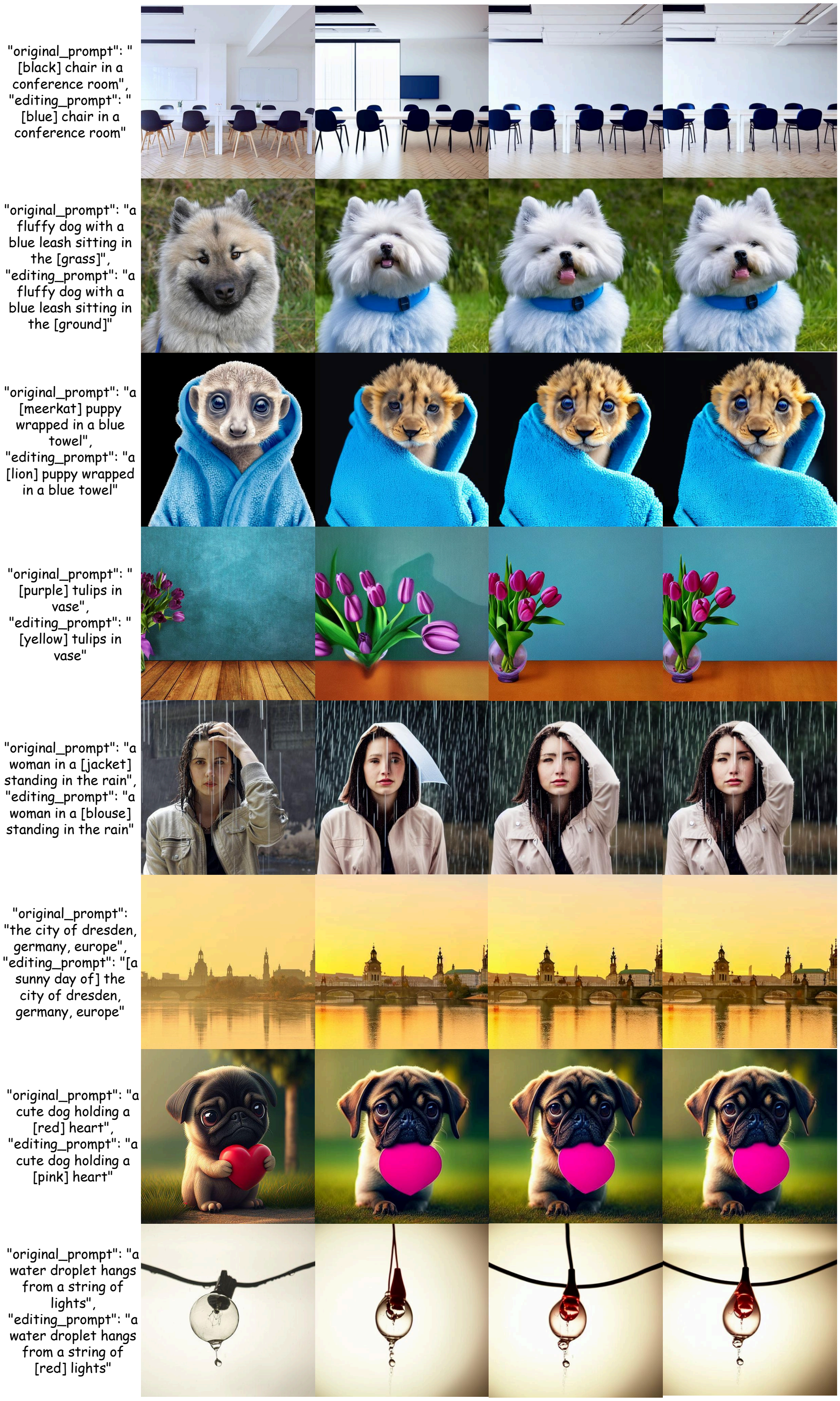
A Appendix

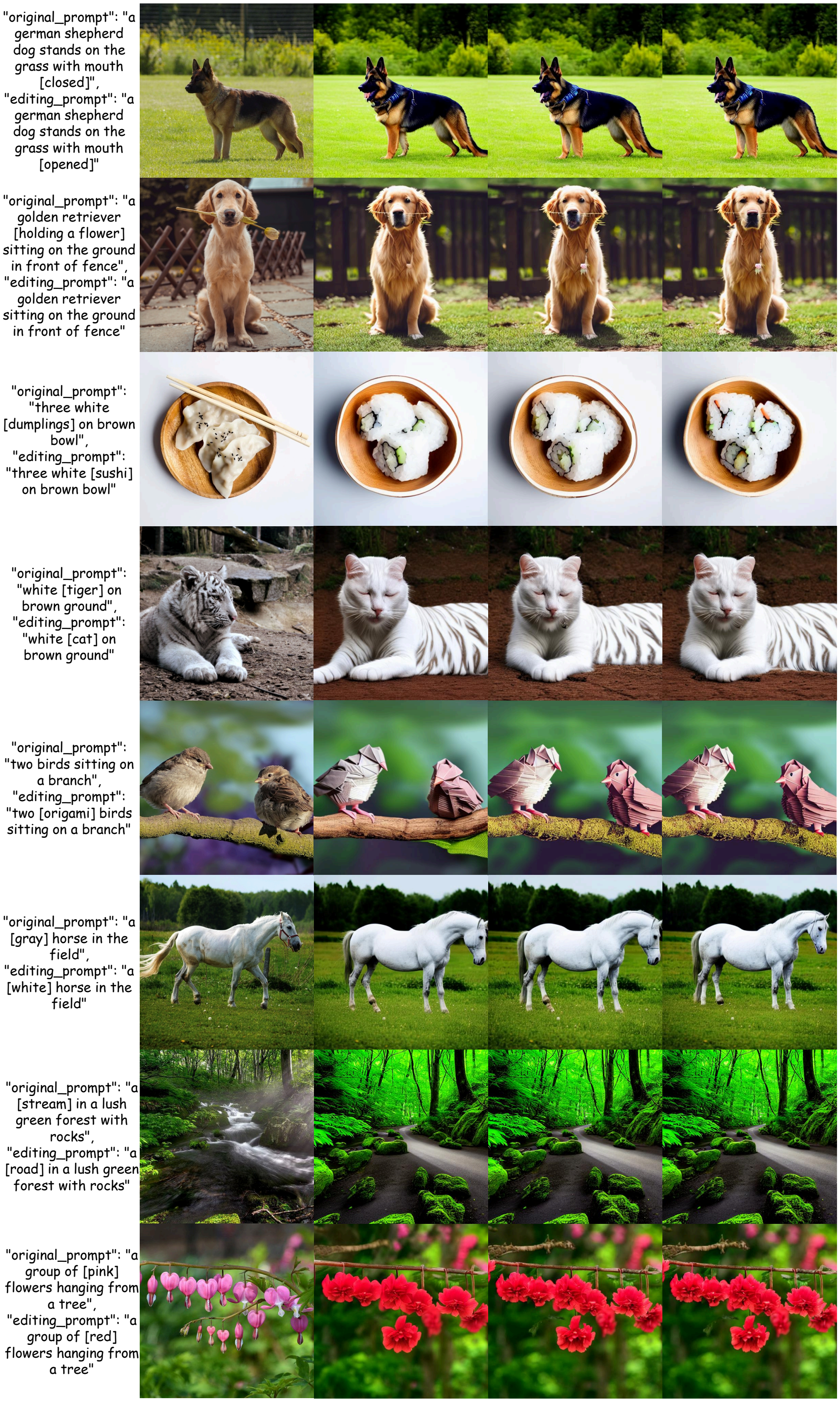
A.1 Loss Trend for 50 steps

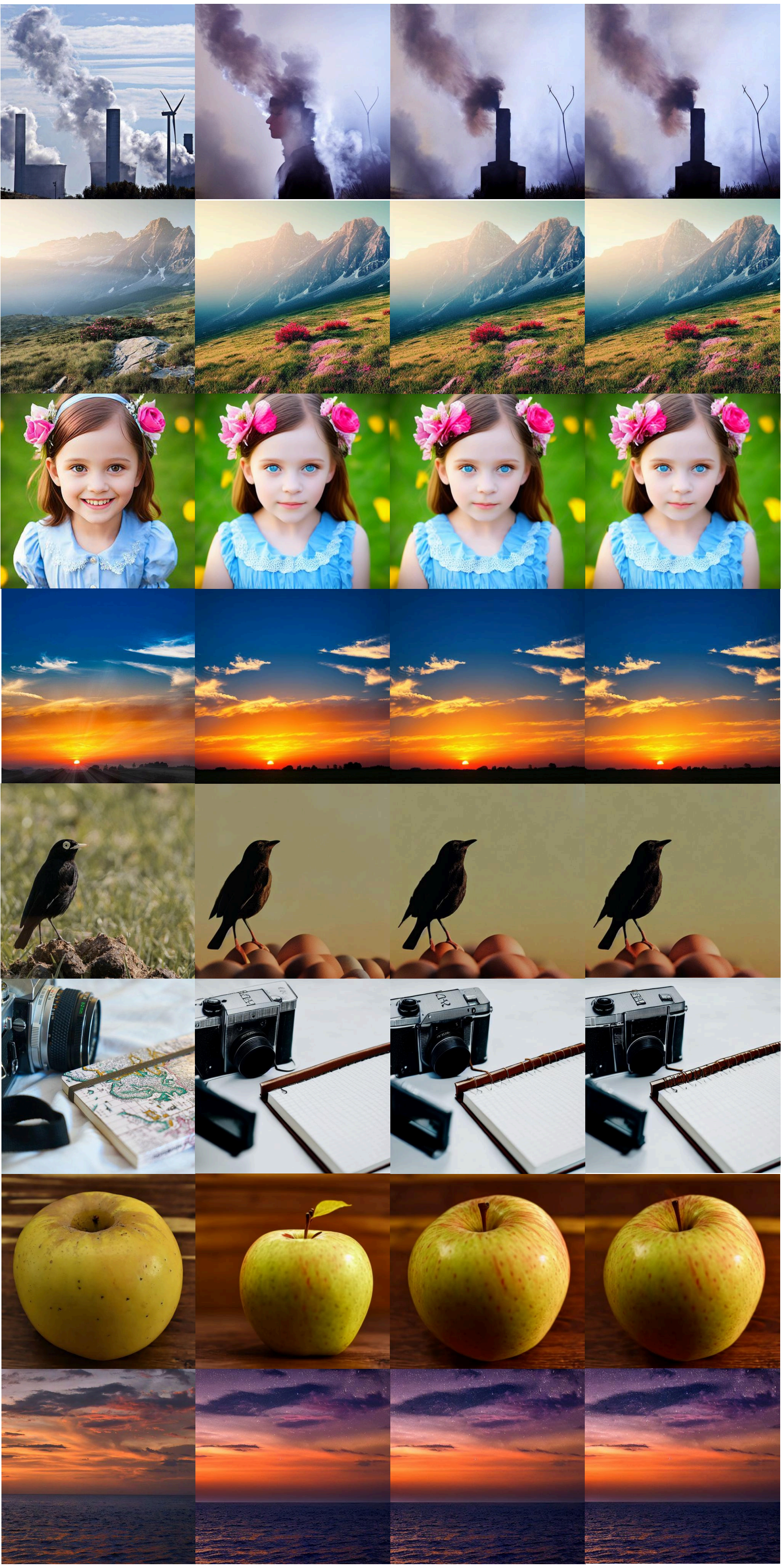
We provide the complete loss trend in Figure 6. It can be observed that the intra path loss exhibits noticeable hysteresis. Since the primary purpose is to demonstrate the hysteresis rather than illustrate the final convergence position, we did not use logarithmic scaling on the coordinate axes.

A.2 More Experiments


Input
P2P
$f(z_t) - z_t$
$f(z_t) - f(\hat{z}_t)$
"original_prompt": "a [round] cake with orange frosting on a wooden plate", "editing_prompt": "a [square] cake with orange frosting on a wooden plate"
"original_prompt": "a [cat] sitting on a wooden chair", "editing_prompt": "a [dog] sitting on a wooden chair"
"original_prompt": "blue light, a black and white [cat] is playing with a flower", "editing_prompt": "blue light, a black and white [dog] is playing with a flower"
"original_prompt": "an [orange] cat sitting on top of a fence", "editing_prompt": "an [black] cat sitting on top of a fence"
"original_prompt": "a [dog] is laying down on a white background", "editing_prompt": "a [lion] is laying down on a white background"
"original_prompt": "meat [balls] on white plate", "editing_prompt": "meat [sushi] on white plate"
"original_prompt": "the statue of liberty holding a [torch]", "editing_prompt": "the statue of liberty holding a [flower]"
"original_prompt": "a [white] horse running in the sunset", "editing_prompt": "a [golden] horse running in the sunset"


"original_prompt": " [black] chair in a conference room", "editing_prompt": " [blue] chair in a conference room"
"original_prompt": "a fluffy dog with a blue leash sitting in the [grass]", "editing_prompt": "a fluffy dog with a blue leash sitting in the [ground]"
"original_prompt": "a [meerkat] puppy wrapped in a blue towel", "editing_prompt": "a [lion] puppy wrapped in a blue towel"
"original_prompt": " [purple] tulips in vase", "editing_prompt": " [yellow] tulips in vase"
"original_prompt": "a woman in a [jacket] standing in the rain", "editing_prompt": "a woman in a [blouse] standing in the rain"
"original_prompt": "the city of dresden, germany, europe", "editing_prompt": "[a sunny day of] the city of dresden, germany, europe"
"original_prompt": "a cute dog holding a [red] heart", "editing_prompt": "a cute dog holding a [pink] heart"
"original_prompt": "a water droplet hangs from a string of lights", "editing_prompt": "a water droplet hangs from a string of [red] lights"


"original_prompt": "a german shepherd dog stands on the grass with mouth [closed]", "editing_prompt": "a german shepherd dog stands on the grass with mouth [opened]"
"original_prompt": "a golden retriever [holding a flower] sitting on the ground in front of fence", "editing_prompt": "a golden retriever sitting on the ground in front of fence"
"original_prompt": "three white [dumplings] on brown bowl", "editing_prompt": "three white [sushi] on brown bowl"
"original_prompt": "white [tiger] on brown ground", "editing_prompt": "white [cat] on brown ground"
"original_prompt": "two birds sitting on a branch", "editing_prompt": "two [origami] birds sitting on a branch"
"original_prompt": "a [gray] horse in the field", "editing_prompt": "a [white] horse in the field"
"original_prompt": "a [stream] in a lush green forest with rocks", "editing_prompt": "a [road] in a lush green forest with rocks"
"original_prompt": "a group of [pink] flowers hanging from a tree", "editing_prompt": "a group of [red] flowers hanging from a tree"



"original_prompt": " [smoke]", "editing_prompt": " [fire]"

"original_prompt": "mountain landscape with [flowers and] withered grass", "editing_prompt": "mountain landscape with withered grass"

"original_prompt": "a little girl with [green] eyes and flowers in a blue dress", "editing_prompt": "a little girl with [blue] eyes and flowers in a blue dress"

"original_prompt": "sunset over a field with clouds and a [bright] sky", "editing_prompt": "sunset over a field with clouds and a [dark] sky"

"original_prompt": "a bird standing on [clods]", "editing_prompt": "a bird standing on [eggs]"

"original_prompt": "a camera and a notebook on a [bed]", "editing_prompt": "a camera and a notebook on a [table]"

"original_prompt": "a [yellow] apple sitting on top of a wooden table", "editing_prompt": "a [red] apple sitting on top of a wooden table"

"original_prompt": "ocean with [clouds] in the sky", "editing_prompt": "ocean with [stars] in the sky"

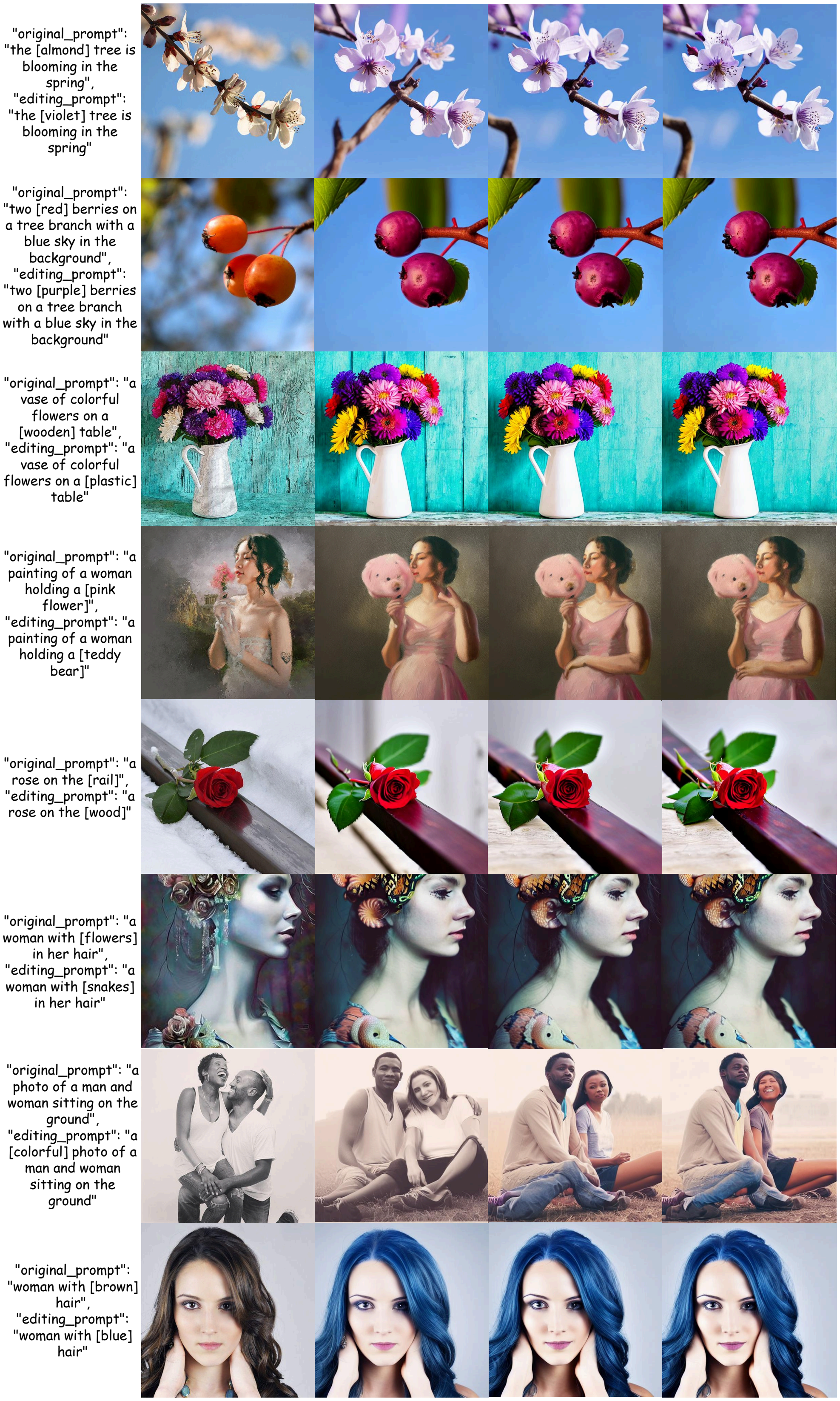

"original_prompt": "the [almond] tree is blooming in the spring", "editing_prompt": "the [violet] tree is blooming in the spring"
"original_prompt": "two [red] berries on a tree branch with a blue sky in the background", "editing_prompt": "two [purple] berries on a tree branch with a blue sky in the background"
"original_prompt": "a vase of colorful flowers on a [wooden] table", "editing_prompt": "a vase of colorful flowers on a [plastic] table"
"original_prompt": "a painting of a woman holding a [pink flower]", "editing_prompt": "a painting of a woman holding a [teddy bear]"
"original_prompt": "a rose on the [rail]", "editing_prompt": "a rose on the [wood]"
"original_prompt": "a woman with [flowers] in her hair", "editing_prompt": "a woman with [snakes] in her hair"
"original_prompt": "a photo of a man and woman sitting on the ground", "editing_prompt": "a [colorful] photo of a man and woman sitting on the ground"
"original_prompt": "woman with [brown] hair", "editing_prompt": "woman with [blue] hair"

Figure 7: The image editing results in PIE-Bench.

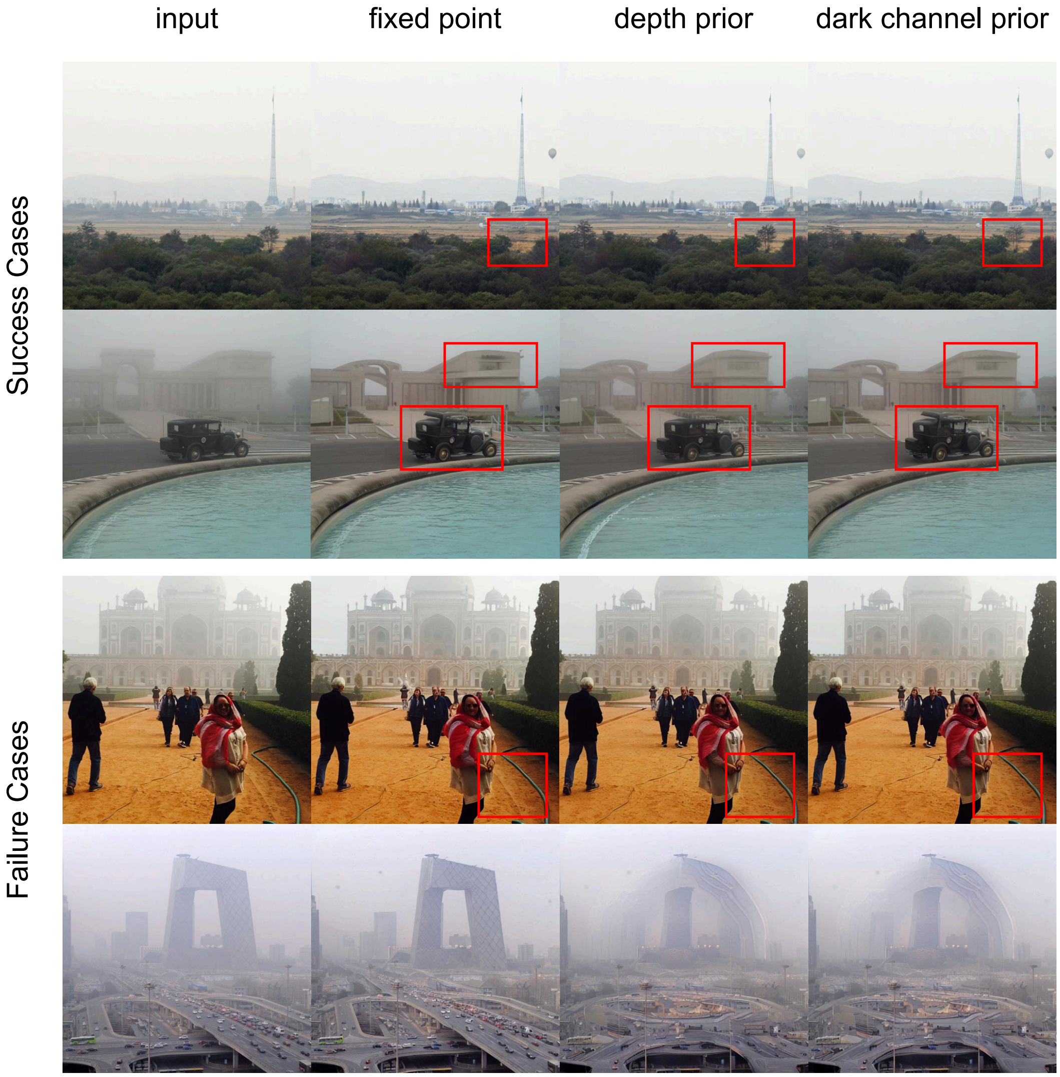


Figure 8: The class-free guidance coefficient is set to 1. The red boxes highlight some noticeable changes, which are beneficial in the success cases but detrimental in the failure cases.

A.3 Dark Channel and Depth Prior

We present success cases and failure cases of incorporating depth prior and dark channel prior, as shown in Figure 8. In the experiments, we directly utilized the fixed point without NTI. To prevent the collapse caused by an increased class-free guidance coefficient, we set the class-free guidance coefficient to 1 in this case.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Listed in the introduction.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See section 7.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The theoretical analysis process can be seen in the main text.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: See section 6.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Our experiment is based on open-source code and data [9].

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See section 6.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Our experiment was conducted according to the benchmark of PIE-Bench [9].

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: The computer resources has already been mentioned [9].

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See section 7.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [No]

Justification: The dataset and reference code are both open-source and free to use for research purposes.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.